A maximum-likelihood framework for determining moving edges in image sequences

Patrick Bouthemy

INRIA

UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France

Tél. (1) 39 63 55 11

Rapports de Recherche

N° 696

A MAXIMUM-LIKELIHOOD FRAMEWORK FOR DETERMINING MOVING EDGES IN IMAGE SEQUENCES

Patrick BOUTHEMY

JUIN 1987

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CÉDEX
FRANCE
Téléphone : 99 36 20 00
Télex : UNIRISA 950 473 F
Télécopie : 99 38 38 32

A MAXIMUM-LIKELIHOOD FRAMEWORK FOR DETERMINING MOVING EDGES IN IMAGE SEQUENCES

UNE APPROCHE PAR VRAISEMBLANCE POUR LA DETERMINATION DE CONTOURS SPATIO-TEMPORELS DANS DES SEQUENCES D'IMAGES

Publication Interne n°365 - Juin 1987
42 Pages

Patrick **BOUTHEMY**

IRISA / INRIA

Campus de Beaulieu

35042 Rennes Cedex

France

Abstract

This report deals with the determination of moving edges in an image sequence. The proposed approach relies on modeling principles and likely hypothesis testing techniques. A spatio-temporal edge in an image sequence is modeled as a surface patch in a 3D spatio-temporal space. A likelihood ratio test enables to achieve its detection and to simultaneously estimate its related attributes. It will be shown that the computation of this test leads to convolve the image sequence with a set of predetermined masks. The emphasis of this report is on a restricted but widely relevant and useful case of surface patch, namely planar one. In addition, we present an implementation of the procedure whose computation cost is merely equivalent to a spatial gradient operator. This method can be of interest for motion analysis schemes not only for supplying spatio-temporal segmentation but also for extracting local motion information. Moreover it can cope with occlusion contours and important displacement magnitude.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(L. A. 227)
UNIVERSITÉ DE RENNES 1 I. N. S. A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(LABORATOIRE DE RENNES)

Résumé

Ce rapport traite de la détermination de contours spatio−temporels dans une séquence d'images. L'approche proposée s'appuie sur des principes de modélisation et met en oeuvre des techniques de tests de vraisemblance. Un élément de contour spatio−temporel dans une séquence d'images est assimilé à une portion de surface dans un espace 3D spatio−temporel. Un test de rapport de vraisemblance permet de réaliser sa détection et d'estimer simultanément ses paramètres associés. Le calcul de ce test revient en fait à convoluer la séquence d'images avec un jeu de masques prédéterminés. Ce rapport s'intéresse particulièrement au cas restreint mais très souvent pertinent et utile du modèle de surface planaire. De plus est présentée une implantation de la procédure correspondante dont le coût calcul est quasiment équivalent à celui d'un opérateur de gradient spatial. Cette méthode trouve son intérêt dans des schémas d'analyse du mouvement non seulement en tant que processus de segmentation spatio−temporel mais également en tant que processus d'extraction d'information locale de mouvement. En outre ce procédé peut prendre en compte des contours dits "d'occlusion" et des amplitudes de déplacement importantes.

Index Terms:

Image sequence, moving edge, local observed velocity, modeling, maximum likelihood test, convolution process, spatio−temporal segmentation, motion analysis.

Table des Matières

Chapitre 1
INTRODUCTION : Problem Statement and Related Work

Image sequence analysis has received more and more attention since 70's. These research efforts have been motivated both by methodological needs and by increasing application domains, [1−2]. Most of these domains are not only interested in spatio−temporal image primitives by themselves. Indeed, these measurements represent intrinsic features conveying information about the environment, i.e. the 3D depicted physical world, [3−6], including relative depth, surface orientation, structure and motion of objects in space, sensor motion.

In the context of dynamic scene analysis, vision understanding schemes must rely on proper and efficient basic tools for spatio−temporal segmentation and motion analysis purposes, [7−8]. This report is concerned with the determination of moving edges in image sequences. More precisely, the designed method simultaneously yields from some local processing the following output concerning moving edges :

- *position;*
- *spatial direction;*
- *velocity component perpendicular to the edge.*

It is well known that only such a partial motion information can be retrieved from a local observation; this problem is often referred to as the *aperture problem*, [9].

The proposed procedure can be perceived as an early processing of some spatio−temporal segmentation scheme and of some motion analysis framework. Segmenting an image according to a spatio−temporal criterion comes to discriminating between fixed and moving areas and between moving areas of different apparent motion. Two approaches can be taken into account : contour−based or region−based ones.

Spatio−temporal segmentation can be an attractive prerequisite process to recover consistent optic flow fields from image sequences. Indeed such a process enables to delimit image domains where a reconstruction procedure, usually comprising some smoothness constraint, like those described in [10−11], can be effectively performed without worrying

about potential discontinuities of the displacement vector field to be estimated. This explicit way to cope with the discontinuity problem can be an alternative to the method proposed in [12], which deals with the so-called *oriented smoothness constraint*. Of course, spatio-temporal segmentation can also represent one fundamental issue to be solved apart from the optical flow reconstruction context. Indeed some of its different aspects can be directly relevant to numerous problems related to dynamic scene analysis, [7-8], such as moving object tracking, [13], robot navigation and obstacle avoidance, [14], or to image coding, such as HDTV transmission, [15].

Up to spatio-temporal segmentation according to a region-based approach, first proposed schemes led to segmenting images into fixed and moving areas, in other words to change detection. Temporal intensity differences have been considered within somewhat heuristic framework in [16]; on the other hand, maximum likelihood tests have been designed in [17], where the spatial gray value distributions were modeled using constant, linear or quadratic approximation. Recently, other approaches achieve to differentiate between several moving regions. They are based on some explicit partial motion information, i.e., the velocity component parallel to the spatial image gradient, which can be extracted from some local processing. Supplementary constraints can be derived for instance by assuming that image regions are projections of 3D planar facets, [18]. In [19], 2D models of motion fields in the image plane are introduced in a hierarchical way and motion-based region segmentation is performed according to maximum likelihood ratio criteria within a split-and-merge procedure.

Of more relevance to this report are other low-level image primitives which are also of key interest for dynamic computer vision, namely spatio-temporal edges (or moving edges). In [20], a simple detector of such primitives relies on the multiplication of the output of a spatial edge gradient operator (e.g. the Sobel one) and the temporal intensity difference between corresponding pixels from two successive images. An extension of the spatial edge detection scheme introduced in [21] is described in [22]. Spatio-temporal edges are given as the zero crossings of the d'Alembertian operator applied to the resultant convolution of the image sequence with a spatio-temporal Gaussian filter.

Moving edge detection skips to moving edge determination when the concern

encompasses the estimation of related features such as spatial orientation and edge–perpendicular component of velocity vector. In [23], space–time relationships, inferred from the way edges move over a tesselation, i.e. a fixed array of cells, enable to determine above–mentioned features. However the detection mechanism is merely assumed to be inherent to these cells without further development. On the other hand, methods designed in [20,22] only or first focus on the detection part. The determination one is somehow postponed. Hence, it must be addressed by another separate technique such as the one primarily developed in [10], and recently thoroughly examined in [24], which is based on the so–called *image flow constraint equation*, relating spatial and temporal derivatives of the intensity function with velocity.

The approach proposed in this report integrates both detection part and determination part within some unified framework. Clearly, this means that, whenever a spatio–temporal edge is detected, its related attributes are simultaneously available. In that sense this method can be of interest for motion analysis schemes not only for achieving segmentation but also for extracting local motion information. More precisely, it can be the initial stage of the estimation of motion fields along contours whose importance has been outlined in [25]. Indeed, this has been implemented in [26].

This report is organized as follows. In Section II, we investigate the modeling of a spatio–temporal edge in an image sequence as a surface patch in a 3D spatio–temporal space. In order to define the detection–estimation criterion, we then express the likelihood ratio which results from the designed hypothesis testing approach. It will be shown that the computation of this test leads to convolve the image sequence with a set of predetermined masks. The emphasis of this report is on a restricted but widely relevant and useful case of surface patch, namely planar one. Section III deals with this case. In addition, we present an implementation of the procedure whose computation cost is merely equivalent to a spatial gradient operator. In Section IV, we give some experimental results obtained both with noisy synthetic data and real images. Finally, Section V contains conclusions. We mention straightforward extensions corresponding to some other tractable modeling cases. We also briefly comment how this scheme is included in a more global motion analysis paradigm we are developing.

Chapitre 2
DETERMINATION OF MOVING EDGES : thereotical analysis

In this section, we present the designed approach for determining moving edges in image sequences. It relies on three basic steps : modeling, hypothesis testing, generalized likelihood ratio definition.

2.1 Modeling of moving edges

Beforehand let us precise some term definitions. We will equally use the expressions 'spatio−temporal edge ' and 'moving edge ', which will be sometimes concisely denoted as *ME* in the remainder of this report. Let us point out that a static spatial edge can also be considered as a spatio−temporal edge whose displacement is nul. The component of the velocity vector perpendicular to the edge will be called 'edge−perpendicular velocity component' or, following [23], 'observed velocity'. It is worth outlining that a ME whose observed velocity is nul can be either a static edge or an edge which is sliding along itself. Finally, as we are concerned with a discrete representation of an image sequence, the notions of displacement vector field and velocity vector field will be confused, although mathematically of different nature.

Let us consider an image sequence as a 3D−space (x,y,t), comprising two spatial dimensions x,y corresponding to every image plane and one temporal dimension t. Then, in this space, a moving edge locally generates a small surface patch. Locally of course must be intended in the three directions x,y and t. Hence the determination of moving edges is stated as the determination of surface patches. The main interest of such an approach is that it can capture both problem aspects : intensity contrast and motion appearance, and moreover from analytical and structural points of view. Besides it contains no a priori inherent restrictions concerning the kind of moving edges to be handled.

Different kinds of surface could be taken into account. For the while, let us only assume that this surface can be defined by some parameter vector Φ. Intuitively, we can argue that this surface results from the interpolation of the different positions (or snapshots), in successive images, of an edge assimilated to a small part of a curve. More formally, if the edge

is supposed to be linear, that-is-to-say a straight-line segment, such a surface is a straight-line-generated one. One relevant particular case corresponds to planar patch. We deal with this example in an extended manner in Section III.

Our choice of such a modeling and the subsequent identification method, described in the next subsection, has been motivated by three main factors. First, the processing to be designed must be *local* in order to get dense measurements (contrary to feature-based matching approach) and to successfully tackle a large range of motion fields. Second, resulting computations must be carried out only at the *intensity array* representation level, if a real-time implementation is aimed. Third, the *underlying edge structure* must yet be somehow integrated in the process (for instance, in order to be able to cope with occlusion configuration).

2.2 Hypothesis testing design

Let us consider an elementary volume π in the 3D space (x,y,t). Two local configurations can be encountered; either there is no spatio-temporal edge inside π or there is one. This can be expressed in geometric and statistical terms as follows. Two competing hypotheses can be acting, respectively :

- H_0 : there is no *ME*; then the intensity distribution within π is modeled as *constant level* c_0 + *noise*, where the noise is assumed to be a zero-mean Gaussian noise with variance σ^2.
- H_1 : there is one *ME*; then a surface patch denoted by $S(\Phi)$ subdivides π into two sub-volumes π_1 and π_2 and the intensity distribution is modeled as previously but according to constant level c_1 within π_1 and c_2 within π_2 with $c_1 \neq c_2$.

Let us recall that Φ represents the geometrical parameter vector which defines the surface patch S; let us assume that Φ has m components, $\Phi=(\Phi_1,..,\Phi_m)$.

The problem now is how to select one hypothesis versus the other one. To this end, likelihood functions corresponding to above-mentioned hypotheses will be considered as described in the next paragraph.

2.3 Likelihood ratio criterion

Let us denote L_0, resp. L_1, the likelihood function associated with hypothesis H_0, resp. H_1. By definition, this function is inferred from the expression of the joint probability density function of the intensities within elementary volume π. As we assume that the intensity random variables for all points within π are independent, the likelihood function is simply given by the product of all density functions. As previously stated, given hypothesis H_0, these density functions correspond to Gaussian laws of mean c_0 and variance σ^2. Hence, the likelihood function L_0 is easily evaluated as follows :

$$L_0 = (2\pi\sigma^2)^{\frac{-n}{2}} \prod_{p_i \in \pi; i=1}^{n} \exp\left\{-\frac{1}{2\sigma^2}(f_{p_i} - c_0)^2\right\} \qquad (1)$$

where $\{f_{p_i}; i = 1, n\}$ are observed intensity values at points p_i from volume π the number of points of which is n.

On the other hand, given hypothesis H_1, the density functions result from a Gaussian distribution with mean c_1, resp. c_2, for all points of sub−volume π_1, resp. π_2, and variance σ^2 in both cases. We then can write the corresponding likelihood function L_1 as follows :

$$L_1 = (2\pi\sigma^2)^{\frac{-(n_1+n_2)}{2}} \prod_{p_j \in \pi_1; j=1}^{n_1} \exp\left\{-\frac{1}{2\sigma^2}(f_{p_j} - c_1)^2\right\} \prod_{p_k \in \pi_2; k=1}^{n_2} \exp\left\{-\frac{1}{2\sigma^2}(f_{p_k} - c_2)^2\right\} \qquad (2)$$

where $\{f_{p_j}; j = 1, n_1\}$ are observed intensity values at points p_j from sub−volume π_1 the number of points of which is n_1 and $\{f_{p_k}; k = 1, n_2\}$ are observed intensity values at points p_k from sub−volume π_2 the number of points of which is n_2. As $\pi = \pi_1 \cup \pi_2$, we have $n = n_1 + n_2$ and $\{f_{p_i}; i = 1, n\} = \{f_{p_j}; j = 1, n_1\} \cup \{f_{p_k}; k = 1, n_2\}$.

Hence, each hypothesis has been expressed in statistical terms by means of its corresponding likelihood function, which must be maximized given the considered hypothesis. To this end, optimal estimators for parameters which this function depends on have to be determined. Then, roughly speaking, the most likely hypothesis is selected by comparing the

likelihood functions evaluated with appropriate optimal estimators. More precisely, we resort to the log−ratio ξ of likelihood functions L_1 and L_0,

$$\xi = \ln(\frac{L_1}{L_0}) \qquad (3).$$

Therefore ξ depends both on geometry parameters Φ, which characterize surface patch S, and on signal parameters c_0, c_1, c_2, which are related to intensity distributions. Let us denote this parameter vector $\Theta = (\Phi_1, .., \Phi_m, c_0, c_1, c_2)$.

Concerning variance parameters, they do not appear as vector Θ−components, since they are assumed to be equal to a constant σ^2. The primary motivation is a practical one. This assumption of constant variance allows to derive a linear criterion as explained in the next subsection unlike a quadratic one otherwise. Moreover it is reasonable from a statistical point of view, as volume π could not have sufficient points for a proper estimation of second−order statistics, and from an image formation point of view as noise variance can be uncorrelated to the signal level. Test robustness is also enforced if less parameters are introduced. The variance σ^2 can be considered as merged into the test threshold λ introduced hereunder. This can justify the eventual use of locally adjusted thresholds.

The resulting criterion for determining moving edges can be stated as follows:

$$\max_{\Phi} \max_{c_1 c_2} \min_{c_0} \xi(\Theta) \geqq \lambda \qquad (4)$$

Clearly, hypothesis H_1 is selected versus hypothesis H_0 if the obtained maximum value for likelihood ratio ξ, i.e. $\xi(\hat{\Theta})$ where $\hat{\Theta}$ denotes the optimal parameter vector, is greater than some predetermined threshold λ. Then, we can conclude that there is a moving edge within considered elementary volume π, whose attributes are precisely given by vector $\hat{\Phi}$ (or as functions of $\hat{\Phi}_1, .., \hat{\Phi}_m$) which has satisfied the above mentioned test in (4).

Yet one problem arises. No analytical closed−forms can be derived to express the optimal estimators $\hat{\Phi}$ corresponding to the geometrical characteristics of the model. In order to point out this difficulty, let us develop the expression of ξ. Substituting in (3) for likelihood function expressions given in (1−2) leads to:

$$\xi(\Theta) = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(f_{p_i} - c_0)^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n_1}(f_{p_j} - c_1)^2 - \frac{1}{2\sigma^2}\sum_{k=1}^{n_2}(f_{p_k} - c_2)^2 \quad (5)$$

Let us recall that sub−volumes π_1 and π_2 are defined once given vector Φ associated with surface patch S. Hence parameters $\Phi_1, .., \Phi_m$ appear in the sum domain boundaries of both last terms in (5), whereas the integrand includes observed intensity values and no formal function.

To settle this question, a set of predefined geometric configurations is considered, $\{\Phi_g, g = 1, .., G\}$. On the other hand, for a given geometry Φ_g, optimal parameters c_0, c_1, c_2 can be expressed. Let us denote $C = (c_0, c_1, c_2)$. They satisfy the following relation:

$$\frac{\partial \xi(\Phi_g, C)}{\partial c_r} = 0 \qquad r = 0, 1, 2 \qquad (6)$$

which leads to

$$\hat{c}_0 = \frac{1}{n}\sum_{i=1}^{n} f_{p_i}; \quad \hat{c}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} f_{p_j}; \quad \hat{c}_2 = \frac{1}{n_2}\sum_{k=1}^{n_2} f_{p_k}. \quad (7)$$

Optimal estimator $\hat{c}_0$, (resp. $\hat{c}_1$, resp $\hat{c}_2$), is merely the empirical mean of the intensity distribution within π, (resp π_1, resp. π_2). $\hat{c}_1$ and $\hat{c}_2$ do depend on given geometric configuration Φ_g, as π_1 and π_2 do. Hence these entities should be denoted as $\pi_{1/g}$, $\pi_{2/g}$ and $\hat{c}_{1/g}$, $\hat{c}_{2/g}$. But this notational precision will not be introduced.

2.4 Convolution−like effectly derived criterion

The following obvious relation can be observed from expressions (7) $n\hat{c}_0 = n_1\hat{c}_1 + n_2\hat{c}_2$. Making use of this relation, starting from (5) and (7) and taking into account simple algebraic developments, likelihood ratio $\xi(\Phi_g, \hat{C})$, or to simplify the writing $\xi(\Phi_g)$, can be expressed in the form:

$$\xi(\Phi_g) = \frac{n_1 n_2}{2(n_1 + n_2)\sigma^2}\ (\hat{c}_1 - \hat{c}_2)^2 \qquad (8)$$

As a matter of fact, it turns out to be interesting to consider the square root ξ of

ratio ξ. Substituting $\hat{c}_1$ and $\hat{c}_2$ for their expressions of (7), we can formulate function ζ as follows:

$$\zeta(\Phi_g) = | \sum_{\eta \in M} a_\eta^g f_{\varrho+\eta} | \qquad (9)$$

where $f_{\varrho+\eta}$ is the observed intensity value at point $\varrho+\eta$, M represents the set of three-dimensional vectorial indices such that $\varrho+\eta$ designates in turn all points belonging to volume π "centered" on point ϱ in (x,y,t) image sequence space. From (7) and (8), we can infer coefficients a_η^g as follows:

$$a_\eta^g = \frac{1}{\sigma} \left(\frac{n_2}{2n_1(n_1+n_2)}\right)^{1/2}, \qquad \text{if } \eta \in \pi_1; \quad (i)$$

$$a_\eta^g = -\frac{1}{\sigma} \left(\frac{n_1}{2n_2(n_1+n_2)}\right)^{1/2}, \qquad \text{if } \eta \in \pi_2. \quad (ii) \qquad (10)$$

Let us precise that the standard deviation σ may not be included in expressions (10i-ii) but rejected in the determination of the test threshold λ as previously explained. Indeed these coefficients only depend on the geometric model features, that-is-to-say the choice of π and Φ_g. Therefore we indicate the index g in the notation of coefficients a's. Accordingly, predefining a geometric configuration Φ_g comes to determining coefficients $\{a_\eta^g, \eta \in M\}$ off-line. In other words, a set of predefined geometric configurations $\{S_g; g = 1,..,G\}$ or $\{\Phi_g; g = 1,..,G\}$ is equivalent to a set of predetermined masks $\{A_g, g = 1,..,G\} := \{a_\eta^g, \eta \in M; g = 1,..,G\}$. Hence, the computation of the ratio $\zeta(\Phi_g)$ has something in common with a convolution operation.

The criterion for determining moving edges can be now expressed as:

$$\max_{\Phi_g} \quad \zeta(\Phi_g) \quad \geqq \lambda \qquad (11)$$

In addition let us outline another aspect of our method. To estimate spatio-temporal features two successive images are sufficient. However, this determination can be sometimes improved if more than two images out of the image sequence are simultaneously considered. For most proposed techniques this extension means troublesome modifications. The formalism developed

in this report can take into account two or several successive images in the same manner. This can be shown by writing function $\xi(\Phi_g)$ as follows:

$$\xi(\Phi_g) = | \sum_{t=t_1}^{t_2} \sum_{\eta \in M_t} a_\eta^g f_{\varrho+\eta} | \qquad (12)$$

where $\cup_{\{t,t_1 \leq t \leq t_2\}} M_t = M$. Subsets M_t correspond to the intersection of the elementary volume π with image planes I_t in number equal to two or greater than two between time t_1 and time t_2.

Chapitre 3
DETERMINATION OF MOVING EDGES : planar patch model case

This section is concerned with a particular but highly relevant ME−modeling example, *the planar patch model case*. It will be stressed that the subsequent implementation of the convolution−like ME−determination criterion then can be, without loss of generality, of a complexity equivalent to a spatial gradient mask operator.

3.1 Geometric considerations

As previously emphasized, the determination of moving edges turns out to the search of surface patches in the (x,y,t)−space. So far, such patches have been assumed to be defined by a parameter vector Φ. Indeed the ME−determination criterion deals with this vector Φ and yields some representative vector $\hat{\Phi}$ if a moving edge is claimed to be present. But, for practical use, ME−attributes must be expressed in terms of spatio−temporal measurements, which have to be derived from surface parameters $\hat{\Phi}$. Hence no gap must appear between the choice of an ideal geometric model and a feasible derivation of effective spatio−temporal measurements.

In the case of a planar patch model, the relationships between surface parameters and spatio−temporal attributes are quite simple. As illustrated in Fig.1, this modeling corresponds to a moving edge stated as a linear one (i.e., a short straight−line segment) which locally undergoes a translation represented by the vector $\underline{V} = (dx/dt, dy/dt, 1) = (\underline{v}, 1)$. Locally must be intended in (x,y)−space, i.e. for the few image points this segment comprises, and in time t, i.e. to the extent of the few considered successive images which the elementary volume π intersects.

The orientation of the planar patch can be described by the following two angles (see Fig.1):

- angle θ with respect to the $x-$axis (with $\theta \in [0, \pi[$);
- angle ψ w.r.t. the $t-$axis (with $\psi \in]-\frac{\pi}{2}, \frac{\pi}{2}[$).

Let ϱ be the reference point of the location of the elementary volume within $(x, y, t)-$space. (ϱ is the point to which the eventually detected ME will be attached). Surface parameter vector Φ is then defined as $\Phi = (\varrho, \theta, \psi)$. Spatio$-$temporal attributes of the moving edge are easily derived. Its location in the image sequence is ϱ; its spatial direction in the image plane is given by angle θ. Its observed velocity or edge$-$perpendicular velocity component is expressed by $v^{\perp} = \tan(\psi)$. Let us notice that the case of a static edge corresponds to $\psi = 0$. It is straightforward to observe that the two components of the velocity vector can not be recovered from the planar patch determination only.

Indeed the planar patch model is the most simple and tractable one. Elements of investigation concerning this representation can be found in [27$-$29]. It is also the most obvious choice when no a priori or contextual information is available, the more so as some local processing is aimed. This planar modeling can be stated as a first$-$order motion approximation, i.e. the local translation component.

Roughly$-$speaking the same holds for the first order development which leads to the image flow constraint equation (which will be concisely denoted as *i.f.c.e.*) relating velocity vector $\underline{v}$, spatial gradient of the image intensity $\underline{\nabla} f$, and its temporal derivative $\partial f / \partial t$ as follows, [24],

$$\underline{v} \,.\, \underline{\nabla} f + \frac{\partial f}{\partial t} = 0.$$

The following advantages can be conceded to the model$-$based approach developed in this report, compared to the gradient$-$based or differential approach. That method presents no *inherent* restrictions concerning both

- kinds of edges likely to be successfully handled (in particular those corresponding to occlusion boundaries), as discontinuities in image distribution are explicitly accounted for;
- extent of measurable motion magnitude, as $\psi \in]-\frac{\pi}{2}, \frac{\pi}{2}[$ implies $v^{\perp} \in]-\infty, \infty[$.

Indeed the same does not hold for the gradient−based methods. Observed velocity fields are often incorrect in the vicinity of occlusion contours, since continuity assumptions required for deriving the i.f.c.e. are no more valid in such areas. Concerning the second assertion, observed velocity can be measured using the i.f.c.e. as long as the local spatial extent of intensity variations and temporal changes interact (this interaction domain can be enlarged owing to prior spatial low−pass filtering). Therefore valid behaviour of the differential approach is usually reached for small displacements. Of course, as far as the model−based approach is concerned, displacement magnitude correctly measurable in practice depends on the image content, especially the edge spatial density. A third point can be outlined. The derivation of the *i.f.c.e.* assumes that changes in image sequences are entirely due to the motion of objects, i.e. that the brightness of a point is invariant in time. This assumption can be relaxed here as long as the edge intensity contrast is not too corrupted by varying illumination, since the measurement of the observed velocity is not directly related to the spatio−temporal intensity changes unlike in the *i.f.c.e.* but is derived from some underlying structure.

3.2 Standard implementation

This subsection is concerned with the computational implementation of the criterion expressed in (11). As previously mentioned and illustrated in Fig.1, $\Phi=(\varrho,\theta,\psi)$ in the planar model case. A given predetermined geometric configuration must be understood as given predefined values for ϱ, θ and ψ. The image which a *ME* will be attached to will be called its reference image.

Concerning parameter ϱ, the set of predefined positions in fact corresponds to the sampling grid, which will be scanned in the reference image. The maximization of ratio $\zeta(\varrho,\theta,\psi)$ with respect to parameter ϱ raises intrinsic difficulties for its implementation. It is not here a matter of *global* maximization, i.e. to find *one characteristic point* in a signal portion like the location of a jump in mean. We are interested in determining a *set* of features in an image which can be close together. Hence the maximization must be *local* in accordance to spatial considerations. However the temporal aspect of the considered features, namely the observed velocity of moving edges, makes the underlying "connection net" for a given *ME* to

enlarge, as exemplified in Fig.2. Then the extent of the maximization domain should be accordingly increased. There is a discrepancy between the two constraints. Some trade−off must be found, which can can be but a sub−optimal implementation of the maximization w.r.t. ϱ. Moreover, whereas efficient procedures are available to fulfil the first constraint, it is much more difficult to comply with the second one because of the inextricable interactions of likely spatio−temporal configurations related to different neighbour points. A maximization procedure will be proposed in the following along with a complementary heuristic to take into account the second constraint.

The notion of geometric parameters will be restricted in the subsequent presentation to $\Omega=(\theta,\psi)$. We consider G possible configurations described by $\Gamma := \{(\theta_g,\psi_g), g=1,G\}$. $\{\theta_g\}$ and $\{\psi_g\}$ are quantized angle values taken within respective definition intervals. Let G_θ, resp. G_ψ, be the number of effectively chosen values for angle θ, resp. for angle ψ; we get $G = G_\theta . G_\psi$. Γ can also be represented as
$\Gamma := \{\Omega_{g_{rq}}=(\theta_r,\psi_q), q=1,G_\psi; r=1,G_\theta\}$.

The next step now is to compute for each considered geometric configuration $\Omega_{g_{rq}}$, or $\Omega_{r,q}$ to simplify the writing, the corresponding mask $A_{r,q}$. In practice, slightly modified expressions are utilized than those reported in (10), to estimate coefficients $\propto_\eta^{rq}$ of mask $A_{r,q}$:

$$\propto_\eta^{rq} = \nu_t . a_\eta^g \qquad (13)$$

where expression of a_η^g is supplied in (10), t refers to the subset M_t which η belongs to as explained after relation (12). ν_t designates a weighting factor to accomodate with the mesh structure induced by the sampling process of the continuous image plane. ν_t is given by $\nu_t = |\varepsilon_1-\varepsilon_2|/(\varepsilon_1+\varepsilon_2)$, where ε_1, resp. ε_2, denotes area of the mesh part belonging to $\pi_1 \cap I_t$, resp. $\pi_2 \cap I_t$. Let Δ_t be the intersection between planar patch $S_{r,q}$ and image plane I_t. As discrete geometry is involved, three cases can be acting as shown in Fig.3. First grid node η may be on line Δ_t, then $\nu_t=0$. Second it may not lie just on line Δ_t but Δ_t still crosses its corresponding mesh, then $0<\nu_t<1$. The third case corresponds to $\nu_t=1$.

Mask shape results from the definition of the elementary volumes π, within which the likelihood−based criterion is computed. In the case at hand, each volume π is a

paralelepiped. More precisely, the intersection of volume π and every image plane I_t is a square of constant size. Hence each three−dimensional mask $A_{r,q}$ subdivises into τ two−dimensional submasks $A^t_{r,q}$ of size $\delta x \delta$ if τ successive images are considered. The positions of the successive submask centers in images I_t depend on parameter ψ_q. More precisely, they are shifted from the position of the submask center in the reference image I_{t_0} at current point ϱ by an interval of $(t - t_0)v_q^{\perp}$, with $v_q^{\perp} = \tan(\psi_q)$, in the direction perpendicular to θ_r.

The implementation of the likelihood ratio criterion for determining moving edges can be described as follows:

ALGORITHM 1

- FOR each point ϱ in the reference image

  FOR each mask $A_{r,q} = \{\alpha_\eta^{rq}, \eta \in M\}$

  CALCULATE $\zeta(\varrho, \Omega_{r,q}) = |\Sigma_{\eta \in M} \alpha_\eta^{rq} f_{\varrho+\eta}|$

  ENDFOR

  SELECT geometry $\Omega_{h,l}$ such that $\zeta(\varrho, \Omega_{h,l})$ is maximum

  IF $\zeta(\varrho, \Omega_{h,l}) \geqq \lambda$ THEN

  MEMORIZE $\varrho, \Omega_{h,l}, \omega(\varrho) = \zeta(\varrho, \Omega_{h,l})$

  ELSE $\omega(\varrho)$ is set to 0

  ENDIF

  ENDFOR

- FOR each point ϱ such that $\omega(\varrho) \neq 0$

  CONSIDER points ϱ_1 and ϱ_2 which lie

  in the direction perpendicular to θ_h on both sides of ϱ

  IF $\omega(\varrho) > \omega(\varrho_1)$ and $\omega(\varrho) > \omega(\varrho_2)$, THEN

  a *ME* is said to be PRESENT at point ϱ whose

  attributes are given by $\hat{\theta} = \theta_h$, $\hat{v}^{\perp} = v_l^{\perp}$

  ENDIF

  ENDFOR

The last part corresponds to the maximization of the criterion with respect to parameter ϱ. It

looks like a thinning process. It only employs local spatial properties; thus it could be insufficient as explained above. An additional heuristic is introduced to avoid false detections and besides to enforce robustness to noise and to high edge spatial density. If $\xi_t(\varrho,\Omega_{r,q})$ denotes $\Sigma_{\eta\in M_t} a_\eta^g f_{\varrho+\eta}$, relation (12) can be rewritten in the form:

$$\xi(\varrho,\Omega_{r,q}) = \left| \xi_{t_0}(\varrho,\Omega_{r,q}) + \sum_{t=t_1}^{t_2\setminus\{t_0\}} \xi_t(\varrho,\Omega_{r,q}) \right| \qquad (14)$$

In fact $\xi_{t_0}(\varrho,\Omega_{r,q})$ does not depend on ψ_q, but outlines the spatial aspect of the *ME*. Then it can be denoted as $\xi_{t_0}(\varrho,\theta_r)$. Before concluding that a *ME* is present at point ϱ according to criterion (11), the following double inequality must be verified for all $t \in [t_1,t_2]\setminus\{t_0\}$:

$$\mu_1 \leqq \left| \xi_t(\varrho,\Omega_{r,q}) / \xi_{t_0}(\varrho,\theta_r) \right| \leqq \mu_2 \qquad (15)$$

where μ_1 and μ_2 are two predetermined thresholds.

3.3 Fast implementation

The implementation as described in ALGORITHM 1 can be quickly time−consuming if a large range of masks is utilized. Therefore a fast version has been designed the complexity order of which is equivalent to conventional spatial gradient convolver. Indeed it corresponds to a particular but highly relevant case, i.e. observed displacement magnitudes $(t-t_0)v_q^\perp$ likely to be directly estimated are limited to entire pixel values. (This implies that subpixel accuracy has to be reached by some local interpolation on the likelihood surface for instance). This assumption induces that mask coefficients α_η^{rq} do not depend any longer on:

1. *time t*; $\forall t, t_1 \leqq t \leqq t_2, v_t = v_{t_0} = v$ (I_{t_0} being the reference image). This is due to the fact that, for a given angle θ_r, every line Δ_t presents the same position with respect to the sample grid.
2. *displacement−related parameter q*; For a given $r = r_e$, submasks $\{A^t_{r_e,q}, q = 1, G_\psi\}$ of the corresponding configuration subclass are all identical. Indeed $v_q^\perp$ only concerns the shift yielding the location of the center of each submask $A^t_{r_e,q}$. Hence, there is no need to compute all the convolution operations as stated in ALGORITHM 1.

As exemplified in Fig.2, the convolution operations at point p arising for instance from the displacement estimation at point p_1 and corresponding to likely configuration $(\theta_r, v_q^{\perp}) = (0, 3)$, or at point p_2 corresponding to likely configuration $(\theta_r, v_q^{\perp}) = (0, -2)$, and so on, or arising as well one image later from the displacement estimation at point p itself for all configurations $\Omega_{0,q}$, are all the same. For the case at hand, two-dimensional submasks $A^t_{r,q}$ are defined by $A^t_{r,q} = A_r = \{\alpha^r_\eta, \eta \in M \cap I\}$, where I designates the image plane. Let $s_{q,t} = ((t - t_0)v^{\perp}_{q,x}, (t - t_0)v^{\perp}_{q,y}, t - t_0)$, where $v^{\perp}_{q,x}$ and $v^{\perp}_{q,y}$ are the $x-$ and $y-$ coordinates of observed velocity. The implementation of the likelihood ratio criterion for determining moving edges is now given by:

ALGORITHM 2

- CONVOLVE each new considered image $I_t = \{f_p\}$ of the image sequence with each mask $A_r, r = 1, G_\theta$

$$A_r * I_t \rightarrow F^r_t = \{\varphi^r_p\}$$

- FOR each point ϱ in the reference image I_{t_0}

  FOR each configuration $\Omega_{r,q}$

  CALCULATE $\zeta(\varrho, \Omega_{r,q}) = \left| \Sigma^{t_2}_{t=t_1} \varphi^r_{\varrho + s_{q,t}} \right|$ (16)

  ENDFOR

  SELECT geometry $\Omega_{h,l}$ such that $\zeta(\varrho, \Omega_{h,l})$ is maximum

  IF $\zeta(\varrho, \Omega_{h,l}) \geqq \lambda$ THEN

  MEMORIZE $\varrho, \Omega_{h,l}, \omega(\varrho) = \zeta(\varrho, \Omega_{h,l})$

  ELSE $\omega(\varrho)$ is set to 0

  ENDIF

ENDFOR

Of course ALGORITHM 2 must be completed by the tinning-like post-processing stage as explained in ALGORITHM 1. One important attractive side-effect of this second version is the following. It does not require that the set of likely velocities $\{v^{\perp}_q\}$ to be predefined at the off-line step of mask coefficient computation. The search area must only be known at the

calculation step of ratio ξ according to relation (16) which merely involves sum over appropriate filtered iintensities.

Chapitre 4
EXPERIMENTAL RESULTS

Before presenting results from experiments with both synthetic and natural images, some complementary remarks will be exposed. A *ME* is obviously beforehand a spatial edge. This spatial aspect can be worthy of being somehow explicitely taken into account. Moreover this would lead to save CPU−time. Let us consider again relation (14)

$$\zeta(\varrho,\Omega_{r,q}) = \left| \xi_{t_0}(\varrho,\theta_r) + \sum_{t=t_1}^{t_2 \setminus \{t_0\}} \xi_t(\varrho,\Omega_{r,q}) \right|$$

One simple relevant constraint can be added; for a given r_e computations corresponding to the evaluation of $\zeta(\varrho,\Omega_{r_e,q})$ for all $q = 1, G_\psi$ are not carried out if

$$\left| \xi_{t_0}(\varrho,\theta_{r_e}) \right| < \gamma . \lambda \qquad (17)$$

with $0<\gamma<1$. Then $\zeta(\varrho,\Omega_{r_e,q})$ is set to 0, for all $q = 1, G_\psi$. This supplementary constraint can also be perceived as a complement to the heuristic (15) or even as an alternative.

4.1 Case of synthetic data

Three sets of examples concerning synthetic images will be reported in this subsection. The mask coefficients which are effectively utilized in all the computer processings which are presented hereunder are in fact the nearest integers to $100.\alpha_\eta^r$.

4.1.1 *Block* example

The first example is a sequence of two computer−generated images depicting blocks, as shown in Fig.4. They differ from a translation of the camera along its axis of view. The intensity distribution is piecewise constant. Four possible spatial directions θ_r have been considered: $\{0°, 45°, 90°, 135°\}$; and 51 observed displacements from $-25 w_\theta \, \underline{n}_{\theta_r}$ to $+25 w_\theta \, \underline{n}_{\theta_r}$ where $\underline{n}_{\theta_r}$ is the unit vector perpendicular to the edge, and w_θ is a metric weighting factor in order to reach entire locations of the grid sample. For instance, $w_\theta = 1$ for $\theta_r = 0°$, $w_\theta = \sqrt{2}$ for $\theta_r = 45°$. Every submask is of size $2x2$. The spatio−temporal edges along

with their estimated observed velocities are plotted in Fig.5. The reference image is the first one. Some false detections only occurs at the bottom part of the greater block; the corresponding boundary in fact disappears from the first image to the second one.

4.1.2 *Polygon* example

The second example includes two computer-generated images of a polygon rotating rigidly in the image plane (Fig.6). The object intensity level is equal to 180, and the background intensity level to 125. A zero-mean Gaussian noise has been added of standard deviation $\sigma = 10$ for the first image and $\sigma = 8$ for the second one. Both images have also been convolved with a set of masks A_r corresponding to the same four values of angle θ as in the previous example. The edge-perpendicular displacement search area raises up to $\pm 30 w_\theta \; \underline{n}_{\theta_r}$. Mask size is $5x5$. These results shown in Fig.7 outline that rotation motion can be successfully handled by this method. Since the considered local modeling assumes that the edge spatial direction does not locally change from one image to the next one, however rotation extent must not be too important.

4.1.3 *Two disks* example

The last example comprises a more critical motion configuration (Fig.8). The foreground disk moves parallel to the axis of view; hence, a dilatation is observed, whose magnitude raises up to 3 pixels along the disk border. Moreover it partially occludes another disk that is shifted by a translation parallel to the image plane of (3,3) down to the lower right image corner. The luminance function l, for each disk of radius r_d and centered on (x_d, y_d), is defined by

$$l(x,y) = L - \frac{L-b}{r_d^2}\left[(x-x_d)^2 + (y-y_d)^2\right] + \mathit{offset}$$

where L is the max gray level within the disk and b the background gray level, to which uniform noise has been added. Quite satisfactory results are shown in Fig.9, with 6 possible values for angle θ, namely $\{0°, 30°, 60°, 90°, 120°, 150°\}$ and a search interval of $[-10, +10]$.

4.2 Case of real images

4.2.1 *Urban scene* example

The first example consists of two successive frames extracted from a natural sequence captured by a TV camera. It represents an urban scene, Fig.10. A panning produces the apparent movement in the image plane, which has been established to be an horizontal displacement of three pixels to the left, apart from some linen and bush areas where other motion interacts. Fig.11 contains results obtained with 6 possible values for angle θ, namely $\{0°, 30°, 60°, 90°, 120°, 150°\}$ and a search interval of $[-5, +5]$. Submask size is $5x5$. Within structured regions such as roof or window borders, estimated *ME*'s are fairly good. On the other hand, highly textured regions are difficult to handle. Fig.12 points out local specific configurations which could lead to erroneous measurements as the search direction is only the perpendicular one.

4.2.2 *Printer* example

The second natural sequence, presented in Fig.13, includes two images of a printer acquired by a CCD camera. Only the printer has been moved from one image to the next, camera and background remain fixed. Fig.14 shows the estimated *ME* field with 4 possible values for angle θ, namely $\{0°, 45°, 90°, 135°\}$ and a search interval of $[-10, +10]$. It can be outlined that the method can successfully cope with rather important displacement magnitude even along occlusion boundaries.

Chapitre 5
CONCLUSION

We have described a model−based approach for determining moving edges which makes use of likelihood techniques. This early processing is relevant to spatio−temporal segmentation paradigms. It can also provide contour−based optic−flow estimation schemes with local motion information. The designed algorithm relies on the local modeling of a moving edge as a surface patch in the (x,y,t) space of the image sequence. This report has dealt in an extensive manner with the case of the planar patch model. The likelihood ratio test is implemented according to a mask convolution procedure, the complexity order of which can be similar to conventional spatial gradient computation owing to certain assumptions but without loss of generality. Results obtained with both synthetic and natural images have been discussed.

More complex modeling could also be handled by this method. If the local spatial 2D−edge is modeled as a circle arc, the surface patch becomes a cylindrical surface patch. Another example could consist in adding a local rotation component to the observed velocity of a 2D linear edge. The resulting surface patch is then of the kind of straight−line−generated surface. This would merely lead to other sets of masks. Thus, an interesting future development could be to study whether this approach could cope with more global configurations as those presented in [30]. Another straightforward extension is the integration of a priori knowledge in the modeling process. For example the knowledge of the camera motion in a static scene is explicitely expressed in the modeling of a moving edge in [31] and the determination of the surface patch parameters allows to completely estimate the velocity vector in this case by making use of the focus of expansion (F.O.E.). In addition, experiments are reported in [31] with more than two successive images considered out of the image sequence for determining moving edges. Let us notice that the method described in this report could be also applied to the problem of edge detection in 3D−imagery and more generally to the determination of $N-1$−dimensional hyperplans in a N−dimensional space.

This model−based likelihood framework for determining moving edges is part of a complete motion estimation scheme from image sequence, which consists of three main stages:

local processing, intermediate-level structuring, optic-flow field estimation, [32]. The first stage is concerned with the determination of spatio-temporal edges as described in this report, and/or the segmentation into regions according to motion-based hierarchically-performed criteria which take into account an explicit partial motion information, [19]. Both designed algorithms utilize principles of local modeling and likely hypothesis testing. The purpose of the second stage is to obtain a structured partition of the image, resulting from edge linking and/or region segmentation. The final stage deals with the velocity field estimation, i.e., the reconstruction of the second component of displacement vectors by combining local observations. First, a recursive stochastic gradient, used to achieve the minimization of some simple functional, enables to estimate optic flow along contours, [26]. Then, the estimation within delineated domains is considered according to the method presented in [11]. Our approach is in particular distinguished by treating beforehand potential discontinuities of the velocity field in the image. Moreover, it produces a set of meaningful intermediate-level spatio-temporal primitives.

ACKNOWLEDGEMENTS

The author would like to thank Albert Benveniste for useful discussions.

This work has been supported by ADI (Agence pour le Développement de l'Informatique) under grant 84/805 project 1707.

REFERENCES

[1] T.S. HUANG, (ed.), *Image Sequence Processing and Dynamic Scene Analysis*, NATO−ASI Series, Vol. F2, Springer−Verlag, 1983

[2] H.−H. NAGEL, "Image sequences −Ten (octal) years− From phenomenology towards a thereotical foundation", *Proc. 8th Int. Conf. on Pattern Recognition*, Paris, Oct. 1986, pp.1174−1185

[3] J. BARON, "A survey of approaches for determining optical flow, environmental layout and egomotion", Univ. of Toronto, Dept. of Computer Science, RBCV−TR−84−5, Nov. 1984

[4] A. MITICHE, "Computation of optical flow and rigid motion", *Proc. 2nd IEEE Workshop on Computer Vision : Representation and Control*, Annapolis, 1984, pp.63−71

[5] A.M. WAXMAN and S. ULLMAN, "Surface structure and three−dimensional motion from image flow kinematics", *The Int. Jal of Robotics Research*, Vol.4, No 3, 1985, pp.72−94

[6] G. ADIV, "Determining three−dimensional motion and structure from optical flow generated by several moving objects", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI−7, No 4, July 1985, pp.384−401

[7] H.−H. NAGEL, "Overview on image sequence analysis", in *Image Sequence Processing and Dynamic Scene Analysis*, T.S Huang (ed.), NATO−ASI Series, Vol. F2, Springer−Verlag, 1983, pp.2−39

[8] R. JAIN, "Dynamic scene analysis", in *Progress in Pattern Recognition 2*, L. Kanal and A. Rosenfeld (eds.), North−Holland, 1985

[9] D.C. MARR AND S. ULLMAN, "Directional selectivity and its use in early visual processing", *Proc. Royal Society Lond.*, Vol. B211, March 1981, pp.151−180

[10] B.K.P. HORN AND B.G. SCHUNCK, "Determining optical flow", *Artificial Intelligence*, Vol. 17, 1981, pp.185−203

[11] A. ROUGEE, B. LEVY AND A.S. WILLSKY, "Optic flow estimation inside a bounded domain", Lab. for Information and Decision Systems, M.I.T., Report LIDS-P-1589, Aug. 1986

[12] H.-H. NAGEL AND W. ENKELMANN, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No 5, Sept. 1986, pp.565-593

[13] G.R. LEGTERS AND T.Y. YOUNG, "A mathematical model for computer image tracking", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-4, No 6, Nov. 1982, pp.583-594

[14] B. ESPIAU AND P. RIVES, "Closed-loop recursive estimation of 3D features for a mobile vision system", *Proc. Conf. IEEE Robotics and Automation*, Raleigh, March 1987

[15] R. STOREY, "HDTV motion adaptive bandwith reduction using DATV", *Proc. Int. EURASIP Workshop on Coding of HDTV*, Vol.2, Italy, Nov. 1986

[16] R. JAIN, W.N. MARTIN AND J.K. AGGARWAL, "Segmentation through the detection of changes due to motion", *Computer Graphics and Image Processing*, Vol. CGIP-11, 1979, pp.13-34

[17] Y.Z. HSU, H.-H. NAGEL AND G. REKERS, "New likelihood test methods for change detection in image sequences", *Computer Vision, Graphics and Image Processing*, Vol. CVGIP-26, 1984, pp.73-106

[18] D.W. MURRAY AND N.S. WILLIAMS, "Detecting the image boundaries between optical flow fields from several moving planar facets", *Pattern Recognition Letters*, Vol. 4, April 1986, pp.87-92

[19] P. BOUTHEMY AND J. SANTILLANA RIVERO, "A hierarchical likelihood approach for region segmentation according to motion-based criteria", to appear in *Proc. 1st Int. Conf. on Computer Vision*, London, June 1987

[20] S.M. HAYNES AND R. JAIN, "Time varying edge detection", *Proc. 6th Int. Conf. on Pattern Recognition*, Munich, Oct. 1982, pp.754–756

[21] E.C. HILDRETH, "The detection of intensity changes by computer and biological vision systems", *Computer Vision, Graphics and Image Processing*, Vol. CVGIP–22, 1983, pp.1–27

[22] B.F. BUXTON AND H. BUXTON, "Computation of optic flow from the motion of edge features in image sequences", *Image and Vision Computing*, Vol. 2, No 2, May 1984, pp.59–75

[23] P. KAHN, "Local determination of a moving contrast edge", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI–7, No 4, July 1985, pp.402–409

[24] B.G. SCHUNCK, "The image flow constraint equation", *Computer Vision, Graphics and Image Processing*, Vol. CVGIP–35, 1986, pp.20–46

[25] E.C. HILDRETH, "Computations underlying the measurement of visual motion", *Artificial Intelligence*, Vol. 23, 1984, pp.309–354

[26] P. BOUTHEMY, "A method of integrating motion information along contours including segmentation", *Proc. 8th Int. Conf. on Pattern Recognition*, Paris, Oct. 1986, pp.651–653

[27] P. BOUTHEMY, "Estimation of edge motion based on local modeling", *SPIE Conf. Computer Vision for Robots*, Vol. 595, Cannes, Dec. 1985, pp.162–169

[28] C. LABIT AND A. BENVENISTE, "Motion estimation in a sequence of television pictures", in *Image Sequence Processing and Dynamic Scene Analysis*, T.S Huang (ed.), NATO–ASI Series, Vol. F2, Springer–Verlag, 1983, pp.292–306

[29] F. GLAZER, "Computing optic flow", *Proc 7th Int. Joint Conf on Artificial Intelligence*, Vancouver, 1981, pp.644–647

[30] A.M. WAXMAN AND K. WOHN, "Contour evolution, neighborhood deformation, and global image flow: planar surfaces in motion", *The*

International Journal of Robotics Research, Vol. 4, No 3, 1985, pp.95–108

[31] L. MARCÉ AND P. BOUTHEMY, "Determination of a depth map from an image sequence", to appear in *Proc. Conf. ICAR*, Paris, Oct. 1987

[32] P. BOUTHEMY, "Détermination du mouvement apparent dans une séquence d'images : extraction de primitives locales, structuration intermédiaire, estimation du champ des vitesses", INRIA–Rennes, Research Report No 619, Feb. 1987

LIST OF FIGURE CAPTIONS

Fig. 1 : Local modeling of a moving edge as a planar patch

Fig. 2 : This diagram illustrates two problems :
a) (§3.2) maximization of likelihood ratio ζ w.r.t. to ρ; P_1 and P_2 are "connected" via P
b) (§3.3) fast implementation of the criterion; Identical convolution operations occur at point P for testing likely configurations $\Omega_{0,3}$ at point P_1 and $\Omega_{0,-2}$ at point P_2

Fig. 3 : Computation of weighting factor $v_t = |\varepsilon_1 - \varepsilon_2|/\varepsilon_1 + \varepsilon_2$ as discrete geometry is involved
* : sampling grid nodes
Δ_T: intersection between planar patch $S_{r,q}$ and image plane I_T
Hatched areas exemplify the measurements of ε_1 and ε_2
Case 1 : $v_t = 0$, case 2 : $0 < v_t < 1$, case 3 : $v_t = 1$

Fig. 4 : "Block" sequence, image size = 256 x 256

Fig. 5 : Estimated ME field for the example of Fig.4. Axes are sampled each five pixels. Every second ME is plotted. $\lambda = 1000$, $\mu_1 = 0.75$, $\mu_2 = 1.25$

Fig. 6 : "Polygon" sequence, image size = 256 x 256

Fig. 7 : Estimated ME field for the example of Fig.6. $\lambda = 4500$, $\mu_1 = 0.66$, $\mu_2 = 1.5$, $\gamma = 0.5$

Fig.8 : "Two-disks" sequence, image size = 256 x 256

Fig. 9 : Estimated ME field for the example of Fig.8. $\lambda = 2500$, $\mu_1 = 0$, $\mu_2 = \infty$ $\gamma = 0.5$

Fig.10 : "Urban scene" sequence, image size = 170 x 447

Fig.11 : Estimated ME field for a part of the example of Fig.10. $\lambda = 2500$, $\mu_1 = 0.8$, $\mu_2 = 1.2$

Fig.12 : Examples of critical situations which may lead to ambiguous or erroneous estimations of $v^{\perp}$
a) corner whose displacement is not colinear to one of its sides ; no corresponding contour points are present in image T_2, for points lying along the hatched part of the corner in image T_1, in the direction perpendicular to the contour. Let us outline that no problem happens if the estimation process is concerned with the displacement from T_2 to T_1
b) step whose displacement is colinear to one of its sides ; false determination may occur for points lying along the hatched part at the bottom of the step.

Fig.13 : "Printer" sequence, image size = 128 x 128

Fig.14 : Estimated ME field for the example of Fig.13. $\lambda = 5000$, $\mu_1 = 0.75$ $\mu_2 = 1.25$

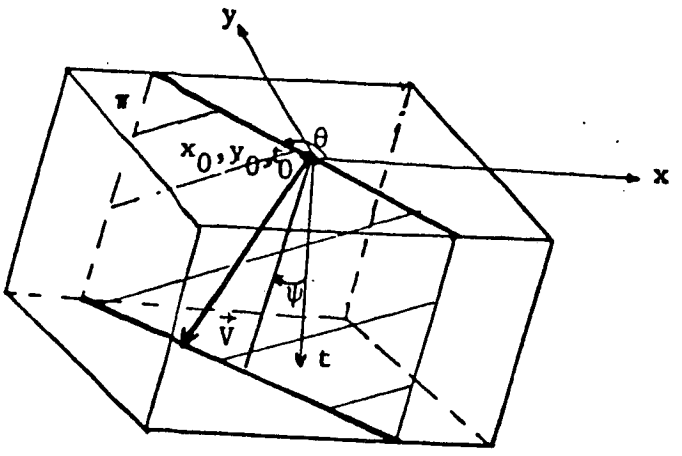

Figure 1

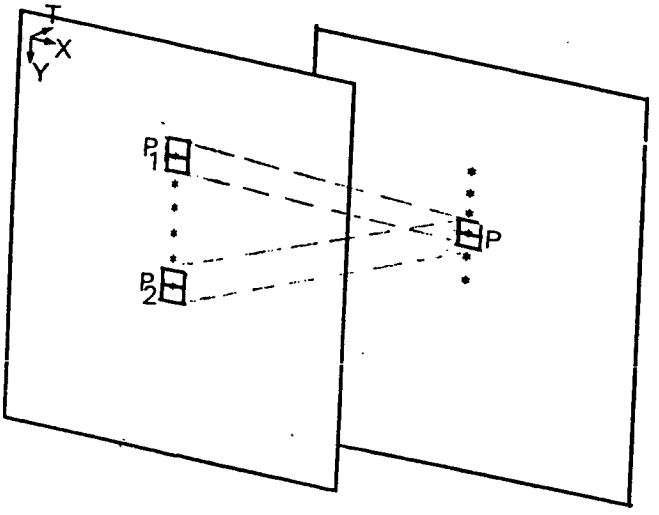

Figure 2

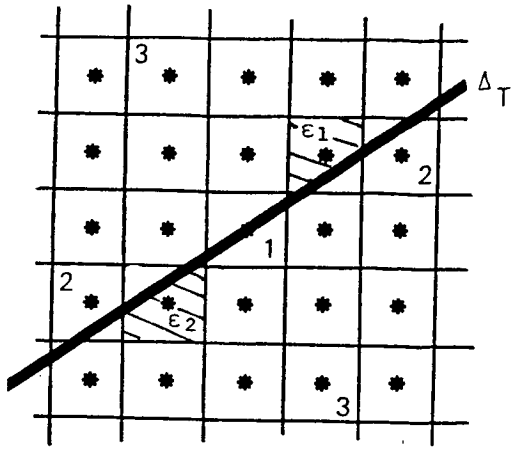

Figure 3

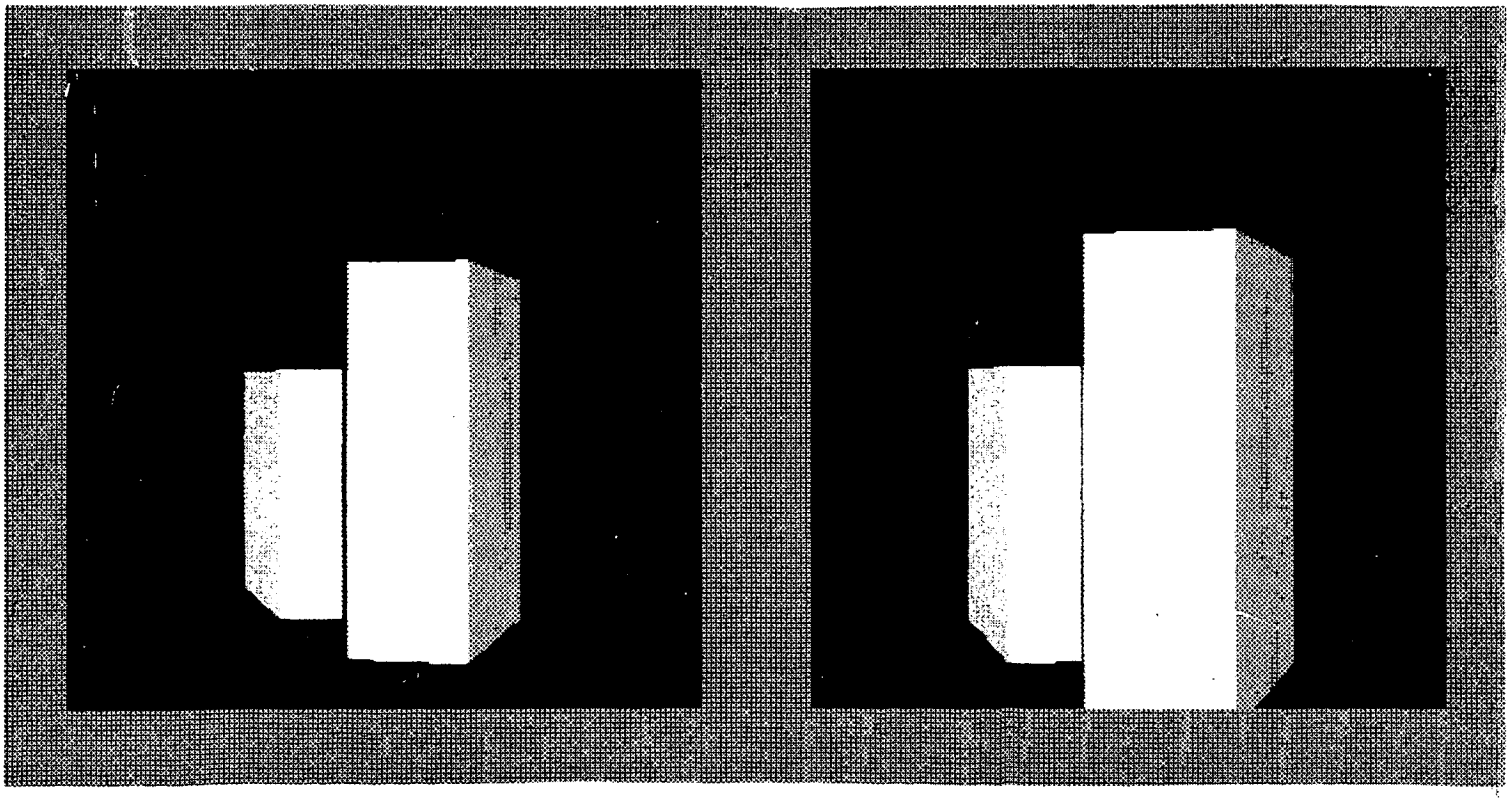

Figure 4

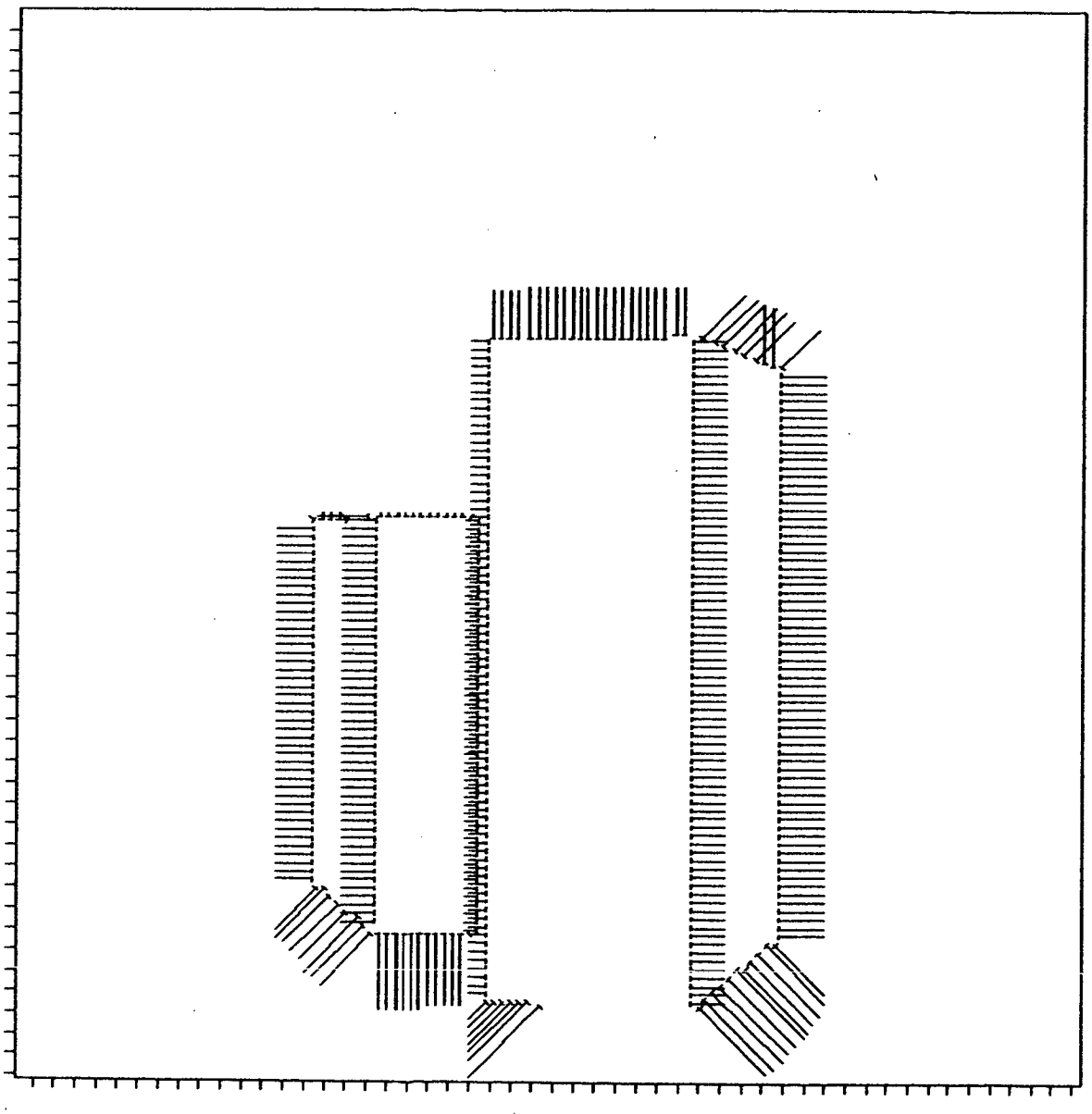

Figure 5

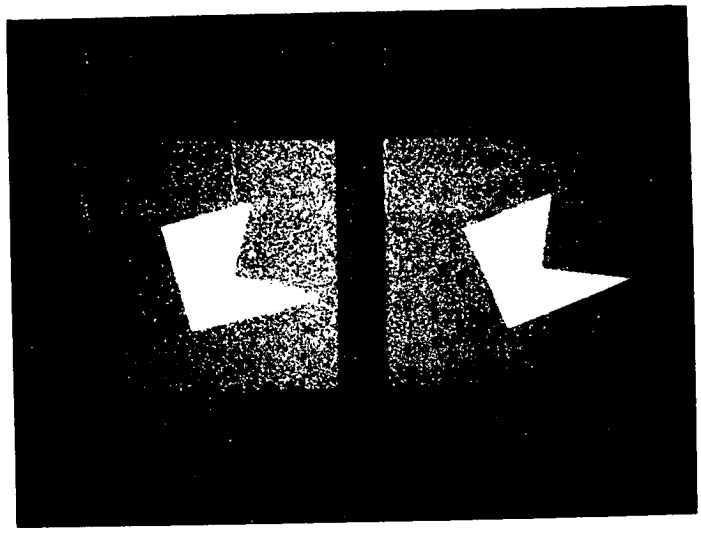

Figure 6

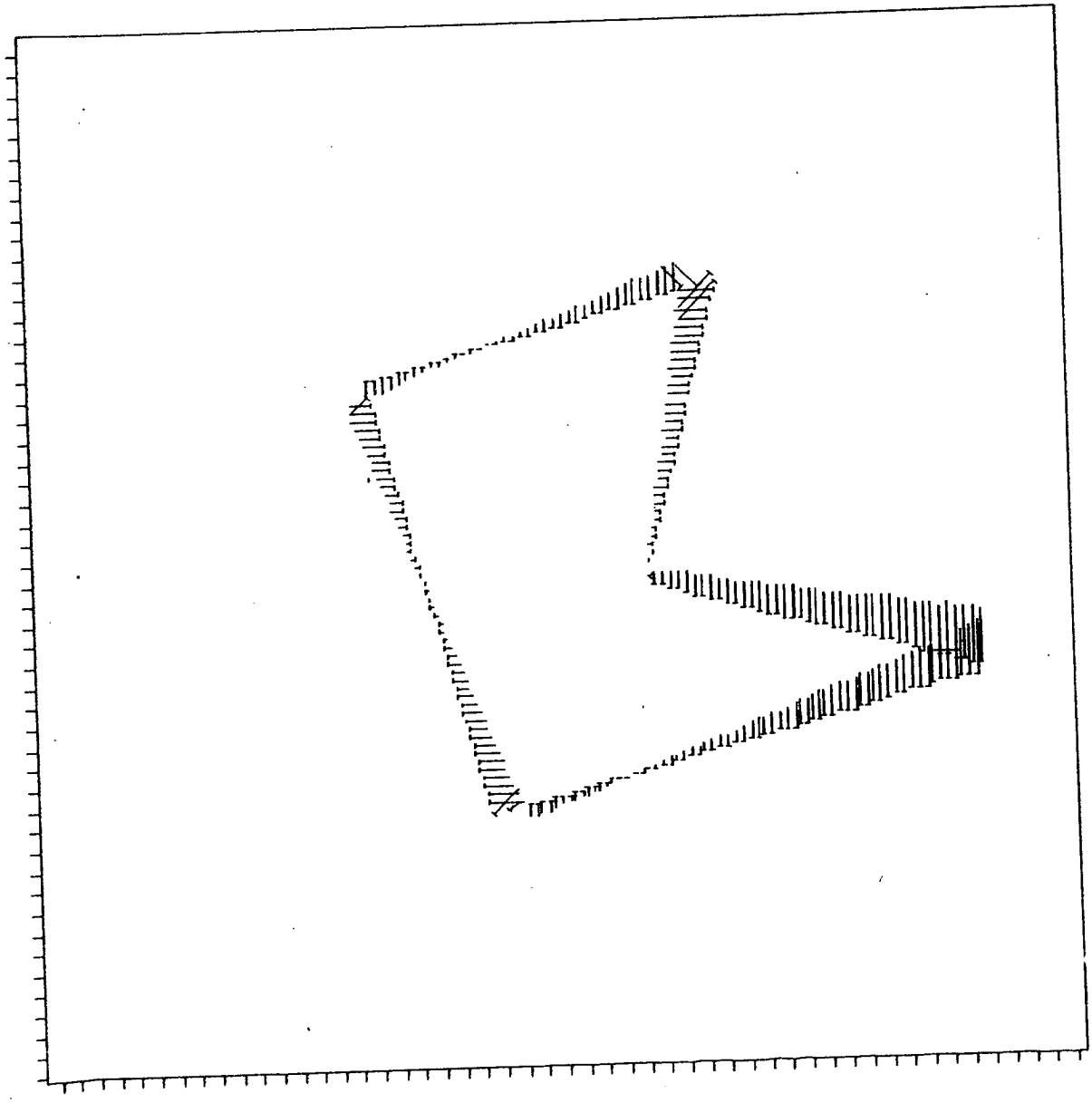

Figure 7

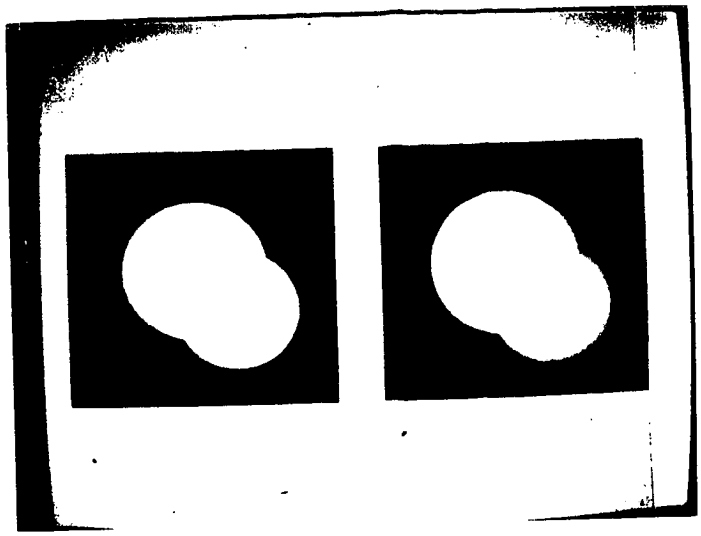

Figure 8

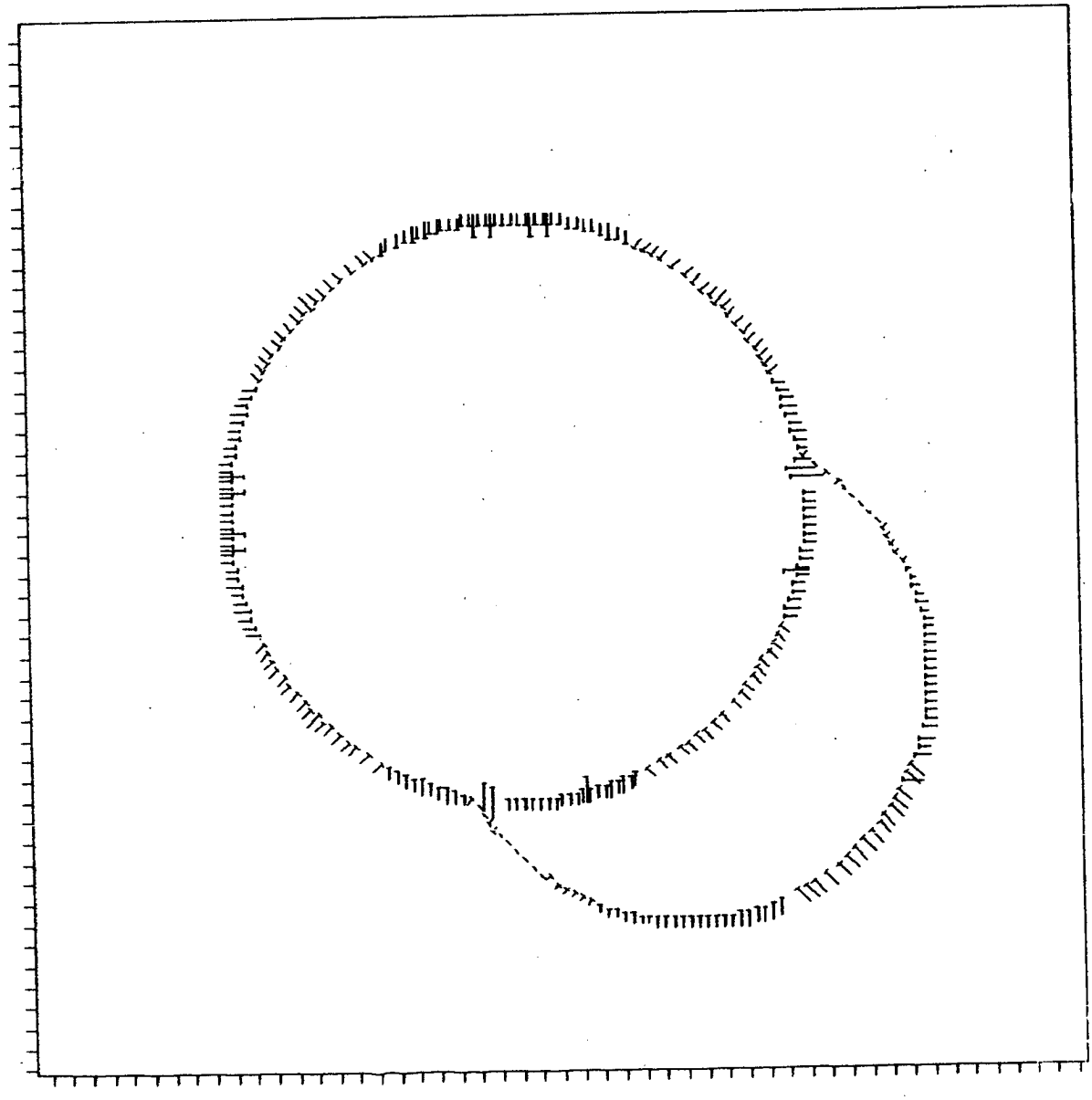

Figure 9

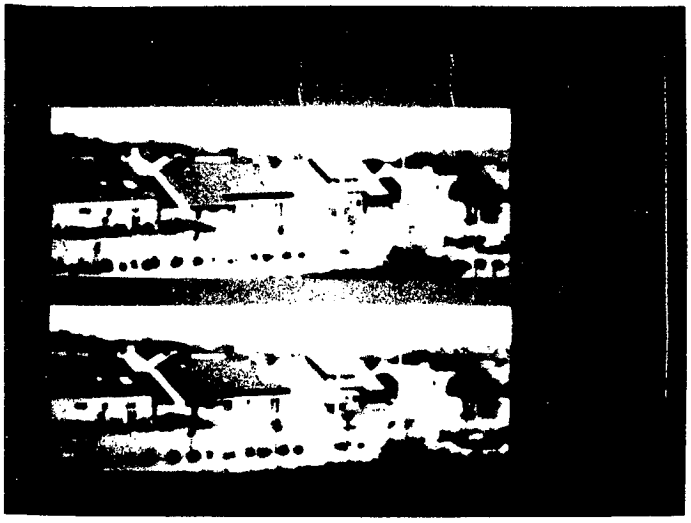

Figure 10

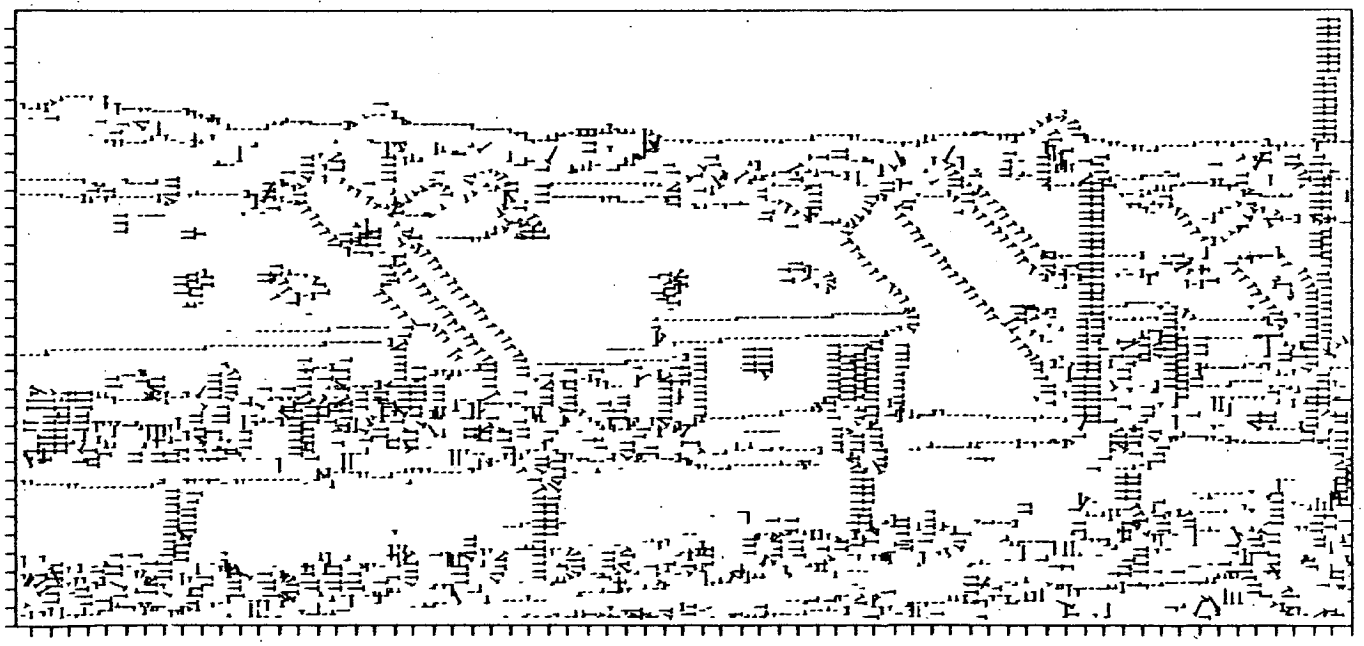

Figure 11

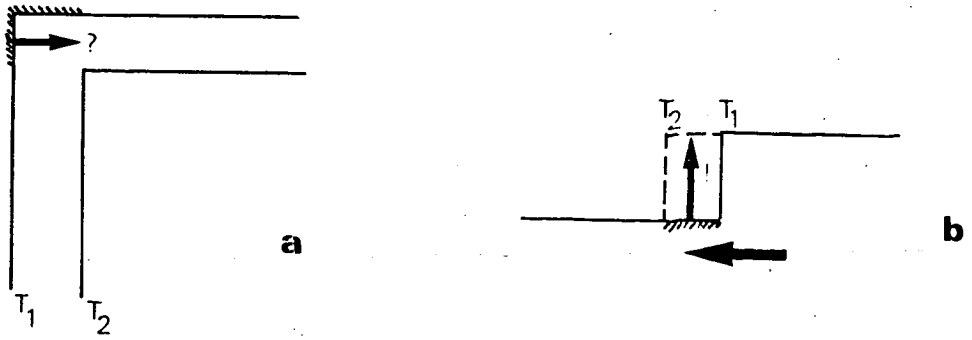

Figure 12

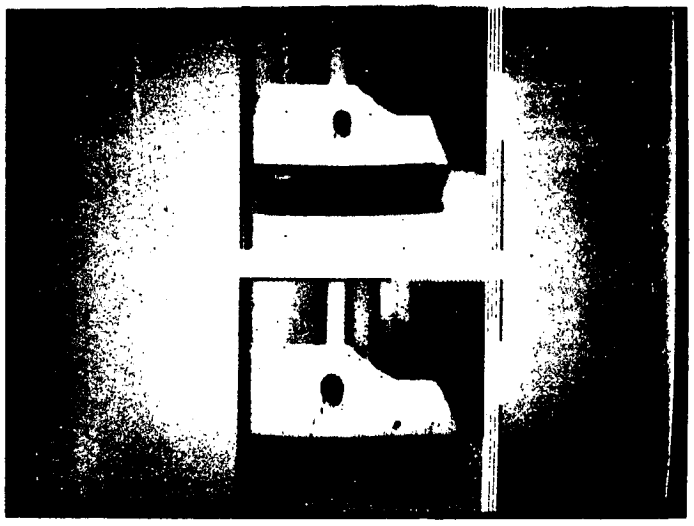

Figure 13

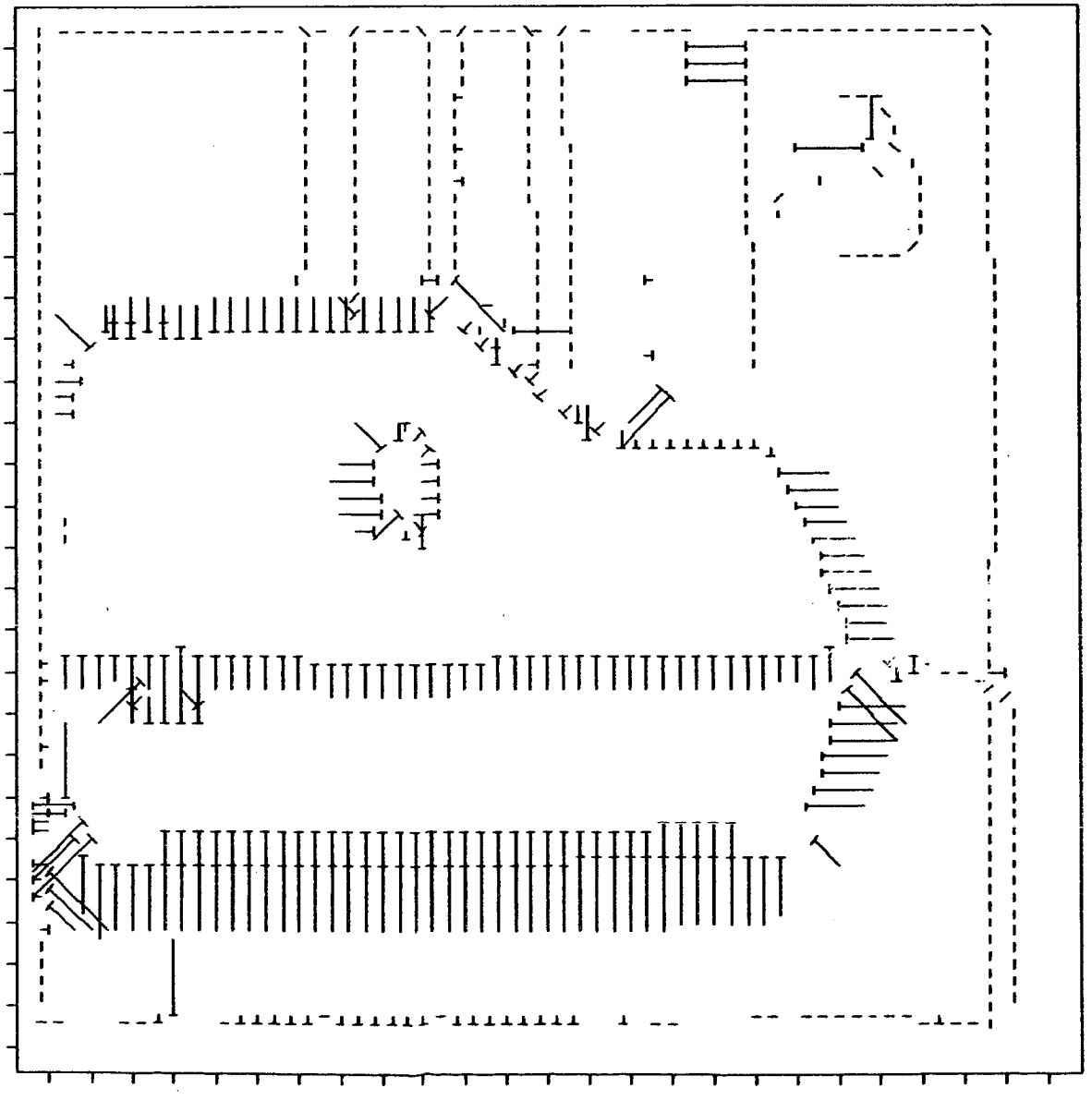

Figure 14

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique